Automatic recognition of company names in speech utterances

The invention relates to a method of automatic recognition of company names in speech utterances. The invention also relates to a dialogue system, more particularly, an inquiry system comprising a processing unit for the automatic recognition of company names in speech utterances.

5   In dialogue or inquiry systems such as, for example, in telephone inquiry systems, the recognition of company names causes particular problems. These problems are based on the fact that users, when pronouncing company names, more often than not do not stick to a predefined fixed format or a certain syntax. For example, parts of a company name are often omitted when entered, abbreviations are formed, acronyms are used or also parts of

10   the company name are exchanged. This leads to unsatisfactory results during the automatic recognition of company names.

Therefore, it is an object of the invention to reduce the error rate during the automatic recognition of company names in speech utterances.

The object is achieved by a method as claimed in claim 1 and a dialogue

15   system as claimed in claim 7.

The recognition results of a speech recognizer customarily used, which still show a high error rate, are subjected to a post-processing according to the invention. For this purpose, a database is used in which all the company names permissible for the respective application are stored. By utilizing the database information, for example ridiculous speech

20   recognition results can be corrected. Also a selection can be made of the best recognition result from many different recognition result alternatives produced by the speech recognizer.

Preferably, the legally correct form of the company names is stored in the database. A word sequence hypothesis produced by the speech recognizer, or a list N of best word sequence hypotheses, respectively, is then compared to the database entries. A search is

25   then made in the database for the word sequence hypotheses as a whole and for parts of the word sequence hypotheses. With the results of the search, a company name stored in the database is now selected as a recognition result while the word sequence hypothesis(es) produced by the speech recognizer is (are) taken into account. If the speech recognizer produces only one word sequence hypothesis for each speech utterance entered, and if no

company name can be found that is completely represented in this word sequence hypothesis, a company name will be selected that contains at least partly the word sequence hypothesis. If the speech recognizer produces various word sequence hypotheses for a speech utterance, the comparison with the database entries is extended accordingly and the best word sequence hypothesis considering the company names stored in the database is determined.

Advantageous embodiments of the invention are defined in the dependent claims.

These and other aspects of the invention are apparent from and will be elucidated with reference to the embodiments described hereinafter.

In the drawings:

Fig. 1   shows a dialogue system connected to the public telephone network and

Fig. 2   shows a processing unit for the automatic recognition of company names in speech utterances, which is used in the dialogue system as shown in Fig. 1.

The dialogue system 1 shown in Fig. 1, or a telephone directory system, respectively, is coupled to a public telephone network (PSTN) 3 via an interface 2, so that a user can access the dialogue system 1 via a telephone terminal 4. A speech utterance of a user can thus be applied via the telephone terminal 4, the public telephone network 3 and the interface 2 to a processing unit 5, which is used for converting speech into text. The processing unit 5 produces a recognition result, which is applied to a dialogue control unit 6, which unit 6 determines, in dependence on the respective application, a suitable speech output to be transmitted to the user. A speech signal to be outputted is generated by a processing unit 7 for conversion of text into speech (for example, speech synthesis unit), while the respective speech output depends on control signals which are transmitted to the processing unit 7 by the dialogue control unit 6.

The processing unit 5 is particularly designed such that company names are recognized with a low error rate. The measures taken here are explained by the block diagram shown in Fig. 2, which diagram shows the embodiment of the processing unit 5. A speech utterance available as an electric signal and coming from the interface 2 is here evaluated by a speech recognizer 10 via a speech recognizer core 11 based on Hidden Markov Models (HMM) while use is made of an acoustic model 12 having acoustic references and a speech model 13. The speech recognizer 10 produces a word sequence hypothesis as a recognition result, which hypothesis contains one or more words describing a company name and, if

necessary, still further words that can be evaluated for the recognition of a company name. Block 14 features a comparing unit which makes a comparison of the word sequence hypothesis produced by the speech recognizer 10 with entries of a database 15. In the database 15 are stored company names which are permissible for the respective application

5   while, more particularly, the legally correct spelling is used. It is then advantageous to remove certain words laid down a priori and hardly contributing or not contributing at all to the distinction of company names (articles, often occurring filling words) from the legal names in the database 15 – and, in consequence, also from the word sequence hypotheses of the speech recognizer 10 – and discard them during the comparison in block 14. The database

10   15 then also contains entries having the respective abbreviated company names, where appropriate, which are then used instead of the unabbreviated company names as a basis for the comparison with a word sequence hypothesis. This is advantageous in that the comparison in block 14 is speeded up, because no evaluation takes place of these filling words. During the comparison in block 14, first a search is made for an entry stored in the

15   database 15, which entry is completely contained in the word sequence hypothesis produced by the speech recognizer 10. If this is the case, this company name is issued as a recognition result 16. If this is not the case, a database entry will be searched for which contains a company name that is contained at least partly in the word sequence hypothesis. Preferably, certain parts of company names are then defined as particularly characteristic and receive, for

20   example, a respectively large weight factor, which is taken into account in the comparison made in block 14. For example, with the company name of Philips GmbH, the part "Philips" will receive a higher weight factor than the part "GmbH". For the company name of "Deutsche Telekom", the part "Telekom" will receive a higher weight factor than the part "Deutsche". Words stored in the database 15, which are defined as words carrying no

25   information that can be used for the recognition of a company name, are used to reduce the word sequence hypothesis by respective parts. Examples for such words are articles, prepositions, filling words and so on.

A search engine used for the comparison in block 14 works in the following way in a preferred embodiment: If a recognition result produced by the speech recognizer 10

30   exactly matches the entry in the database 15, this entry receives the highest score – other database entries, which only partly match, may then be issued as alternatives. Preferably, the speech recognizer 10 produces not only one word sequence hypothesis, but a plurality of N best word sequence hypotheses (N > 1) for a speech utterance. These hypotheses are sorted in accordance with a probability determined by the speech recognizer 10, which is taken into

account by the comparing unit 14. More particularly, for one speech utterance not only N best word sequence hypotheses are applied to the comparing unit 14, but, furthermore, a probability value for each word sequence hypothesis, while speech recognizers are used that deliver N best recognition results with respective probabilities P_{Nbest} (company name). Also

5   the comparing unit 14, by evaluating the entries of the database 15, produces probabilities $P_{comparing\ unit}$ (*company name*) for each company name found. The weighting of the final search results may then be made, for example, via the overall probability:

P(company name) = P_{Nbest} (company name) * $P_{comparing\ unit}$ (company name)

This is effected under the simplified assumption that the speech recognition and the

10   comparing unit are statistically independent.

Block 19 features the determining of the speech model values of the speech model 13. In a training phase the entries of the database 15 are evaluated for this purpose. Improvements for the construction of the speech model 13 are achieved in that variants of the company names (block 17) stored in the database 15 defined a priori as appropriate (for

15   example, appropriate mix-ups of parts of company names, colloquial formulations such as, for example, "Big Blue", and others) are entered during the training of the speech model 13. A further improvement of the training of the speech model 13 is obtained in that data, which were recovered from actual inquiries or dialogues, respectively, by means of dialogue systems already in use, are also entered during the training (these data are featured by block

20   18). They can be entered in two ways: on the one hand, in that they are simply added to the training material and, on the other hand, in that the frequencies contained therein of inquiries for certain companies are entered as weight factors (in the sense of a unigram) into the training material which consists of the pure database entries. Furthermore, an on-line adaptation of the speech model 13 is provided with the present speech recognizer 10, which

25   adaptation leads to a further reduction of the error rate for the recognition of entered company names. For the on-line adaptation are used the word sequence hypotheses recovered by the speech recognizer 10 during the operation of the dialogue system. The algorithms for speech model adaptation are known, as are the algorithms for speech model training and these algorithms are combined in block 19.